DESIDOC Journal of Library & Information Technology, Vol. 31, No. 4, July 2011, pp. 226-233

Social Semantic Digital Library: The Future

Projes Roy and Dipti Arora*

Shaheed Rajguru College, Jhilmil Colony, Vivek Vihar, Delhi-110 095

**Defence Scientific Information & Documentation Centre, Delhi-110 054*
E-mail: dipti30arora@gmail.com

ABSTRACT

Conversion of classic and digital libraries into digital libraries is now passe. With evolution of technologies, classic libraries have upgraded themselves and have emerged as library and information centres. This paper discusses the integration of computer technologies with library and information science, and how semantics can be customised for library operations, and what can be the measures for converting routine tasks to semantic tasks. The paper concludes that making social semantic digital library is not at and the only thing required is to reduce the gap between library and Semantic Web communities.

Keywords: Ontology, Semantic Web, social semantic digital library, WWW consortium, Web technologies

1. INTRODUCTION

Digits, digital, digitisation, ontology, metadata, Semantic Webs were the terms developed in computer science but intruding the libraries. Pushing libraries and information centres to cyberage, and serving the key ingredient, i.e., information to the right user at the right time, these terms and technologies have played a vital role. Digitisation and metadata have secured enough place in libraries and information science, but the terms semantics and ontology are comparatively new, and still need to be adopted. Semantic Web and ontology are terms that are often used interchangeably. In computer science the word ontology stand for "representation of entities, ideas, and events, along with their properties and relations, according to a system of categories"[1]. Ontologies provide structural framework for organising information, whereas Semantic Web is the format or the technology that enables machines to understand human readable data. Semantic enables formal organisation of information.

Since the beginning of computers, formal computer languages have been categorised as low-level languages and high-level languages. Low-level languages are machine languages, i.e., the actual 0110100 language and high-level languages are the languages which human can understand and discuss. Semantics is something which correlates between these low-level and high-level languages. It emerged as a separate field under the existing world wide web (WWW) or Web of data. As most of the information on the Web is designed explicitly for human consumption, still people working on Internet carry different views on WWW. Some people say that they could directly get the information which they had intended to search. Some opine that they are happy with the Web as designed for human browsing; they get more or less what they want, and are also finding other things of interest as they browse.

Over the years, the Web technologies have also undergone several changes and advances as per the requirements and the availability. Web 2.0 is one such technology that came into existence with the advent of such technology upgradation. As users of WWW observed drastic changes with the way material was flowing on the web, they were forced to call advanced Web as Web 2.0. Features like personal websites, content management system, message boards, buddy lists, address books, etc., were left for the earlier version, i.e., web 1.0, and the new terms like blogs, wikis, Google personalised, Google scholar, book search, community portal, online social networks were categorised as web 2.0.

What we want, what we see, and what we get has lot of other things attached to it, and every style carries

different meanings. This meaning and the meaning of the meaning has been utilised and evolved into new technologies that forced mankind to think that new generation of Web, i.e., Web 3.0 is emerging. Semantics, as by its nature, is now one of the important components of Web 3.0. It is providing the existing Web a logical view to provide a consistent readability. The way technologies have been upgraded, LIS also has upgraded. A lot of changes are still expected. Semantics Webs need to be adopted with the feeling that we should not think what Semantic Webs can do for us, but what we can do with the Semantic Webs.

2. THE VISION

Bush gave the idea of the 'memex' machine in the 1940, which was based on a universal library, complete with a searchable catalogue. Current Web is more or less influenced by this 'memex' machine and provides a vision for Semantic Web from the current web.

Tim Berners-Lee, the original man behind the WWW, envisioned the Web as repository of richer descriptions of documents and links between these. However, in an effort to provide a simple, usable, and robust working system, these ideas were put to one side, and the simpler, more human-mediated Web, which we know today, resulted.

Tim Berners-Lee, Jim Hendler and Ora Lassila have provided a compelling vision of a world where people need not put much labour to retrieve information through the Web but the Web should come forward directly to carry out routine tasks such as scheduling appointments, finding documents and locating services, giving added information, and giving links to the added information, i.e., the Web itself can do the hard work for them. Their vision more or less falls under artificial intelligence and/or fuzzy logic, where machine itself will be forced to think and act as a human brain. This can be done by providing sufficient context about resources on the Web and also providing the tools to use the context so that machines (or 'software agents'-programs working on behalf of people) can find the right things and make decisions. In the words of the article[2]:

"The Semantic Web brings structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users. Current Web focuses on documents whereas Semantic Web focuses on creating a Web of data. This new approach will allow machines and automatic processes to access information and provide reasons about the (intended) meaning of documents.

According to Gartner[3], an information technology research and advisory company, "During the next ten years, web-based technologies will improve the ability to embed semantic structures in documents, and create structured vocabularies and ontologies to define terms, concepts and relationships. This will offer extraordinary advances in the visibility and exploitation of information, especially in the ability of systems to interpret documents and infer meaning without human intervention. The grand vision of the Semantic Web will occur in multiple evolutionary steps, and small-scale initiatives are often the best starting points."

According to Frank van Harmelen[4] states, "Semantic Web technology is like Nikasil (coating in the cylinders of a car). Very few car drivers are aware of it, but they are aware of reduced fuel consumptions and the extended lifetime of the engine. Semantic Web technology is the Nikasil of the next generation of human-friendly computer applications that are being developed right now. People can only feel the change in the performance and output of their search from Internet, but could not guess how and where the Web changed the search mechanism."

3. SEMANTICS IN KNOWLEDGE ORGANISATION

Broadly speaking, development of Web and also the emergence of Semantic Web from the existing Web moves around Cutter's objectives. Almost a century ago, i.e,. much before the development of the Web, Cutter gave some objectives for cataloguing. In the book, Cutter's Rule for Dictionary Catalogue, he gave some principles to meet those objective that even influence current cataloguing activities. He said, "Subject specificity is expressed by informing cataloguers to enter a work under the subject heading of a class which includes that subject." Most call numbers in the Cutter classification follow conventions offering clues to the book's subject. The first line represents the subject, the second the author (and perhaps title), and the third and fourth dates of editions, indications of translations, and critical works on particular books or authors. Size of volumes is indicated by points (.), pluses (+), or slashes (/ or //).

For some subjects, a numerical geographical sub-division follows the classification letters on the first line. For example, 1 stands for history and criticism of a subject, 2 for a bibliography, 5 for a dictionary, 6 for an atlas or maps, 7 for a periodical, 8 for a society or university publication, and 9 for a collection of works by different authors.

Cutter's rules are simple and are close to the keywords, which one can try to search the book. The same logic is needed to be provided to the current Web, i.e., semantics+current Web. The Cutter's rule that were described specially for subject classification in the

library science are applicable to WWW and computer science also. As the WWW is already facing the information overloading, and also the way information is dispersed through various media, forgetting even about the useful and useless information, we need to apply Cutter's principle to shed the information overloading, and also to make the right information available to the right user at the right time even for WWW.

With the complexity of the availability of information, vast resources that are accumulated as a backdrop of daily activities, there is a bulk of information available on Internet. Such information needs organisation to make it ready for use. The Semantic Web is creating such infrastructure for knowledge acquisition, knowledge representation, and knowledge utilisation. Marshall and Shipman[5] gave three perspectives to sort out the objectives of the Semantic Web (Fig. 1). They said that on one axis we can keep the representations used on the Web, as moving from the particular to the universal." Here particular means which are limited to the authors original motivation for publishing something on the web, and the universal means which are useful in any context. On the other axis, we can keep kept the user, using any of the representations. The users may be either human, who use or retrieve the information directly, using any of the queries or as a result of interacting with a web application, or it may be a machine that do computational processes, either meshing up the information holdings of specific applications or may be weaving a silent mesh of knowledge through the work of agents.

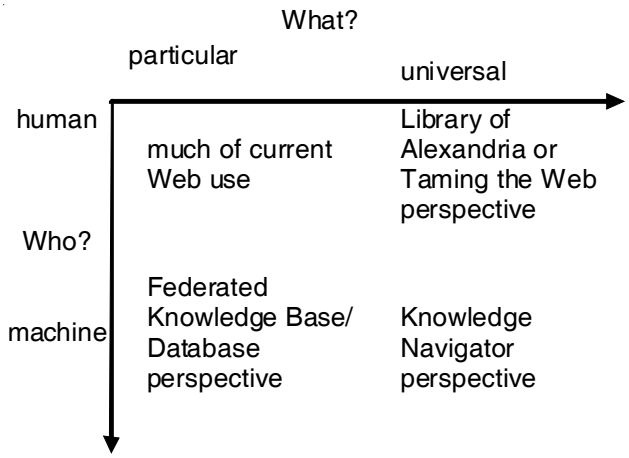

Figure 1. Perspectives of semantic webs.

4. EVOLUTION OF LIBRARIES

Libraries have always been engaged in providing better services to the users with maximum utilisation of their resources and available technologies. Starting with just storage of books, these are now well established and are using every new computer technology and advancement to provide quality services to their patrons. With meager available resources and few services to now open access, open libraries all libraries have come of age. During the last century, libraries have undergone dramatic changes—starting from classic libraries—to scientific libraries—to digital libraries. Twenty-first century will see the evolution of semantic libraries. Semantic digital library is a digital libraries+Semantic Web+social networking (Web 2.0). It will integrate classic digital libraries, with Semantic Web and social networking, and other Web 2.0 technologies. It will integrate information based on different metadata, e.g. resources, user profiles, book marks, and taxonomies on either metadata or communication level, or both, delivering more robust, user-friendly and adaptable search and browser interfaces empowered by Semantic Web technologies, expressiveness of annotations and interoperability with other services (not only digital libraries). The Web 2.0 approach allows users to be engaged in the annotation, and knowledge-sharing process, and thus make semantic digital libraries more useable. Schematic of formation of a semantic digital library is shown in Fig. 2.

5. SEMANTICS CUSTOMISED FOR LIBRARY SERVICES

The WWW itself is more or less like a digital library with digital content and metadata as key commodity. Though many libraries have also upgraded themselves as digital libraries and open libraries available on internet, still if a mixture of semantic is added, these will be more user-oriented and less labour-intensive. It is the need of the hour to add and mix semantics to digital libraries as both share many similarities. Some of these are[6]:

- Abundance of information and material availability in a variety of forms
- Better services, full information access, and knowledge discovery are the mission statements for both, libraries and Semantic Web.
- International and national standards, which help library and the Semantic Web grow more and more.
- Information and resource sharing and collaborative spirit, which has advanced both libraries and Semantic Web.

Based on the similarities, Semantic Web can be used in libraries in a number of ways. Libraries and information centres perform some basic tasks such as knowledge acquisitions, i.e., building the collection; cataloguing; reference; and circulation. All these tasks with their corresponding similarities with semantics are shown in Fig. 3.

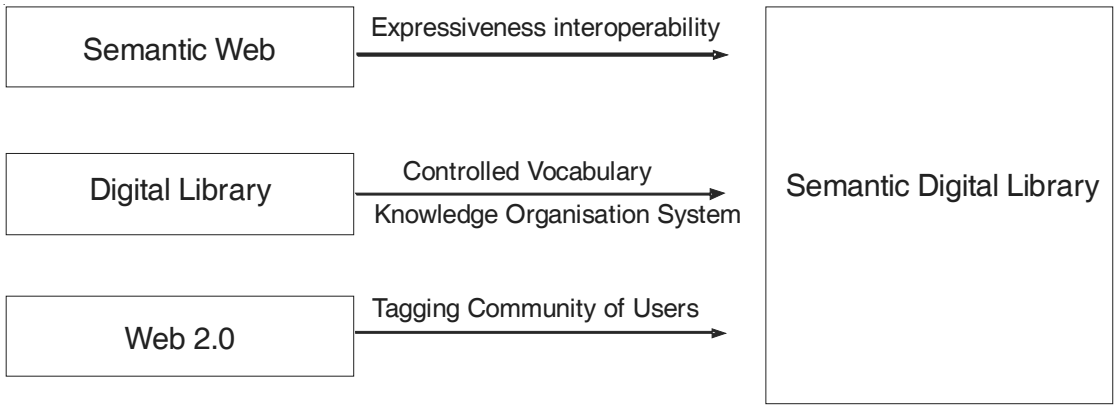

Figure 2. Formation of semantic digital libraries.

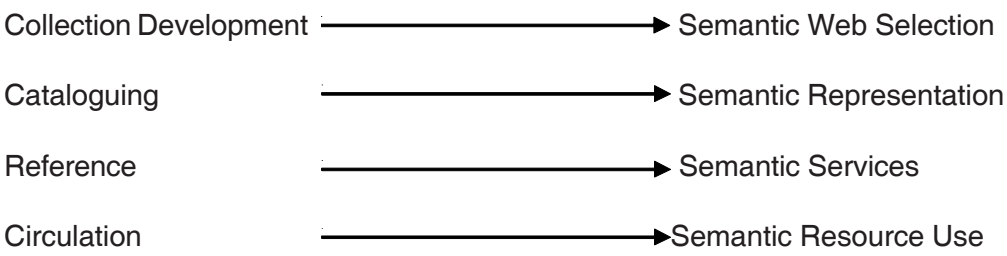

Figure 3. Making routine tasks to semantic tasks.

5.1 Collection Development

Collection development aims to build and develop a coherent collection to serve its respective users. A contract, that is actually a written collection development policy, is maintained between the library and its users. It also includes guideline and selection criteria to acquaint the users how to acquire knowledge in various formats. Libraries and information centres need to review and revise these guidelines and policies as per the present demand and the user change. As new areas of studies and also disciplines are emerging, libraries should also change their agreements to serve at the best possible levels. Correlating the collection development, though Semantic Web selection is not oriented for the specific type of user or community, some specific Semantic Web projects are there that serve specific type of population, topics, or areas.

The http://musicbrainz.org/[7] is one such website or Web project that is a user-maintained open community which collects, music metadata in the form of a relational database and makes it available to the public. Figure 4 shows the home page of the website of musicbrainz.

There are a number of library collection development guidelines and resources available like Guidelines for Writing Collection Development Policies. One can use these resources for developing Semantic Web selection policies. With these master guidelines, one can also develope future projects.

Figure 4. Home page of the musicbrainz.

5.2 Cataloguing

Cataloguing allows library collection findable and easily retrievable even if the user is not professionally sound. The user may know the name of the book, or author, or title, publisher, etc. The best cataloguing facilitates the users with the material available even with least knowledge about the same.

Since both, Semantic Web and cataloguing, deal with the representation of the knowledge, the similarities between the two are obvious. McGuinness[8] found the fine boundary between the representations standards

between the two environments, i.e., Semantic Web and the library cataloguing. He said, "How should a Semantic Web project decide which metadata schema or ontology to use: What level of representation is required to properly represent the information entity so that an agent can successfully manipulate the information and provide a useful service?"

As cataloguing revolves around its objectives, principles and standards like MARC and CCF to manage standardised policies around the globe, similar are the various technologies and formats for semantics Web like resource description framework (RDF), Extensible Mark up language (XML), and web ontology language (OWL). If proper documentation of such technologies with globally accessible standards and Semantic Web representation policies are created then a general framework may also be developed to assist future projects feeding into the overall Semantic Web initiative.

5.3 Reference

Reference service is establishing the contact between the reader and the information through personal service. Reference services keep some of the documents ready for providing immediate services. These may include dictionaries, encyclopedias, bibliographies. yearbooks, directories, standard work of information such as gazetteers, etc. Though the scenario changed after the arrival of computers and electronic resources and slanted more towards keeping the users abreast of current developments and advances with their current awareness services, they (users) were enhanced to personal interaction, dissemination of documentation, and educative and outreach activities. One of the extensions of reference service is community outreach. To meet the basic objective of serving the users, the library and information centres may adopt community outreach.

Similar to the reference services policies, if standardised metadata standards are followed, then it will be easier for Semantic Web algorithms to track the data. Reference and User Services Association (RUSA)[9] has recommended certain guidelines for the information services which address information services from the following perspective (i) Services, (ii) resources, (iii) access, (iv) personnel, (v) evaluation, and (vi) ethics

To serve the users best, Semantic Web projects can refer to the RUSA's Guideline for Information Services. Tables 1 and 2 show service policy and selection policy statements respectively from RUSA's Guidelines for Information Services and demonstrate how the statements could be modified to guild Semantic Web development.

5.4 Circulation

There was a time when users had to wait for the arrival of their copy or have to book it in advance. But the digital technologies have changed the whole scenario, and multiple users can access the same source of information at the same time, apart from the other benefits also. Using digital libraries' objectives and principles, it is

Table 1. Library reference and semantic web project service policy statements

Library Reference Service Statements*	Semantic Web Project Service (SWPS) Statements
RUSA Statement 1.1	**SWPS Statement 1**
The goal of information services is to provide the information sought by the user. Information service should anticipate as well as meet user needs. It should encourage user awareness of the potential of information resources to fulfill individual information needs.	The goal of Semantic Web project *x* is to provide the information sought by agents (computer and human). The Semantic Web service should anticipate agent needs. It should facilitate agent awareness of the potential the services' its information resources can fulfill to individual needs.
RUSA Statement 1.3	**SWPS Statement 2**
The library should strive to provide users with complete, accurate answers to information queries regardless of the complexity of those queries.	Semantic Web project *x* should strive to provide agents (computer and human) with complete, accurate answers to information queries regardless of the complexity of those queries.
RUSA Statement 1.6	**SWPS Statement 3**
The library should actively publicize the scope, nature, and availability of the information services it offers. It should employ those media most effective in reaching its entire clientele or selected segments of that clientele, as appropriate.	Semantic Web project *x* should actively publicize the scope, nature, and availability of the information services it offers. It should employ those media most effective in reaching its entire clientele or selected segments of that clientele, as appropriate.
RUSA Statement 1.7	**SWPS Statement 3**
The library should survey and assess the information needs of its community and create local information products to fulfill those needs not met by existing materials.	Semantic Web project *x* should survey and assess the information needs of its community and create local information products to fulfill those needs not met by existing materials.

Table 2. Library reference and semantic web project evaluation policy statements

Library Reference Service Statements*	Semantic Web Project Evaluation (SWPE) Statements
RUSA Statement 5.1	**SWPE Statement 1**
The library should regularly evaluate its information services to ensure that the service furthers the institution's goals and that the goals reflect the needs and interests of the community served.	Semantic Web project *x* should regularly evaluate its information services to ensure that the service furthers the initiatives goals and that the goals reflect the needs and interests of the community served.
RUSA Statement 5.2	**SWPE Statement 2**
The library should integrate the perspectives of staff and community in the overall evaluation procedure for information service.	Semantic Web project *x* should integrate the perspectives of managers, staff, and user agents (computer and human) in the overall evaluation procedure for Semantic Web service.

**RUSA's Guideline for Information Services (2000).*

easy to develop the Semantic Web policies and guidelines for circulation services. Though many libraries are adopting the open access policies, still some of these have made mandatory the subscription, sometimes paid, or may be just login-password-based. Semantic Web resource use policy can be generated for computer and human agents. For example for accessing, an identification number (digital signature) can be used; for tasks like loan period and renewing loans, policy can be formed to indicate how long an agent can access the project resources and how access privileges can be renewed. Policies for other events, as to provide a venue for agents to request additional resources or semantics; to indicate explain the arrangements of resources within a project, or to inform agents any tracking or recording data use can also be formed.

6. ARE PRESENT LIBRARIES SUFFICIENT?

The way the libraries and information centres have adopted the technologies and have come up with new system, format, standards, and upgradation in their services, prove that the present libraries and information centres are prepared and ready to accept and customise the technologies. Categorising to the traditional types like classic libraries, scientific libraries, digial libraries, etc., they have really come up with amazing features and services. Apart from serving the user to the best possible extent, these are finely mixing the 4Ps of marketing mix (product, price, place, promotion). Marketing mix helps to position the library and information services firmly in the perceptions of their communities served—the wider community for the academic library or the clients and customers and to reach target audience meeting specified objective[10]. The Semantic Web services are quite compatible with libraries services. If both are used together, it will be wonderful cohesion of technology with information.

7. TAMING THE OVERHEADS

As information overloading is not only the problem of information centres but also of WWW, the WWW need to face the same challenges. Some of the challenges are[11]:

- *Vastness*: The World Wide Web contains at least 26 billion pages (March 2011). People keep on uploading the information; sometimes without even deleting the earlier version, they upload the fresh version and fresh webpages. The website links and the webpages contained and maintained even by a single website are enormous. So the Semantic Web policies need to face this challenge. An automated reasoning system can be adopted to overcome this overhead.
- *Vagueness*: Some time, information uploaded on the websites is vague in the context that it is not relevant or may be of no use for any one. It happened hardly that sometimes any particular page is accessed, or may be vaguely people use the words irrespective of their meaning. For that, Semantic Web may use additional thinking, i.e. fuzzy logic technique, to deal with such vagueness.
- *Uncertainty*: If precise things or values are required, then uncertain things and values carry no importance, rather these are futile. Semantics of similar words may match and may be used at different places for different meanings. Then, if a user seeks a particular phrase or word, he may get uncertain results. Probabilistic reasoning techniques can generally be employed to deal with such uncertainty issues.
- *Inconsistency*: Since standards may vary while uploading, or logical contradictions may be there cause of large ontologies, combined from separate sources. It may give inconsistency to the world wide web. In Semantic Webs, one can use Deductive logical reasoning to overcome such inconsistencies.

World Wide Web Consortium (W3C) Incubator Group for Uncertainty Reasoning for the World Wide Web (URW3-XG) Final Report combines these problems together under the single heading of "uncertainty". Many

of the techniques mentioned here will require extensions to the OWL.

8. FUTURE OF SOCIAL SEMANTIC DIGITAL LIBRARY

Best utilisation of resources and technology in Semantic Web made WWW as virtual digital library . At the click of a button, one can get desired result with all thick and thin possibilities. But, this is not the end, we need to form communities and also be more social in sharing knowledge and information. The new community that will arrive will be social semantic. Technologies of Semantic Web like RDF may set a framework to model any kind of metadata and can deliver it to certain level of technologically interoperability. Figure 5 shows how the concept has been evolved from libraries to social semantic digital library[12]. The social semantic digital library will help digital library to build heterogeneous networks of Semantic Web. It may deliver more robust, user-friendly, adaptable search and browsing interfaces empowered by semantics.

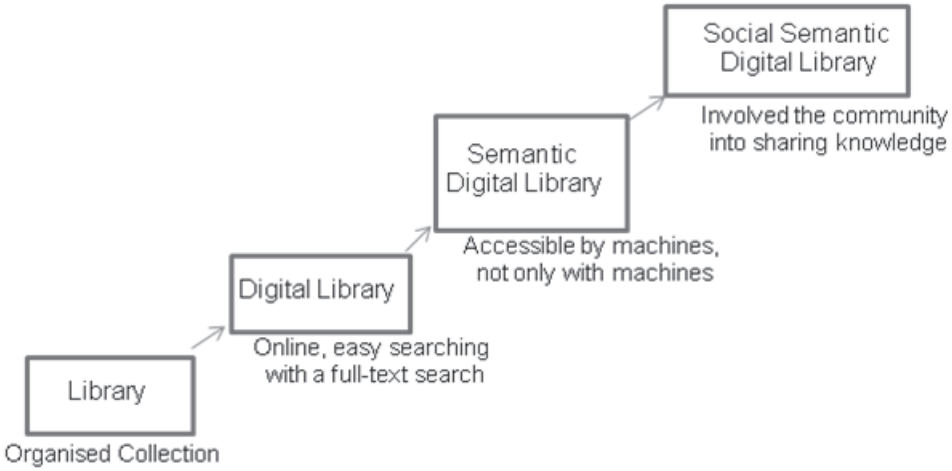

Figure 5. Heading towards social semantic digital library.

9. CONCLUSION

Semantic Web digital library would contain features like semantic blogs, semantic wikis, semantic search, social semantic digital libraries, semantic social networks, semantic social information spaces, etc. These will have open access, open information, and open source. Semantic Web initiative is not difficult for libraries, but information professionals need to acquaint themselves with wide technologies and standards (RDF, OWL, etc.), so that they can generate compatibility of semantic with their vast knowledge for faster and user-based information dissemination. Though there are some other factors like communication barriers, absence of metadata representations, absence of user-friendly applications, limited available literature to constraints the librarian to initiate and adopt the semantics Web, but librarians need to participate in ontologies and semantic-based conferences to explore the technology more and also to give wider coverage to their skills and talent.

REFERENCES

1. Ontology, http://www.wikipedia.org/wiki/Ontology_information_science (accessed on 14 March 2011).

2. Matthews, Brian. Semantic Web technologies. World Wide Web Consortium, JISC.

3. http://www.gartner.com/technology/home.jsp (accessed on 14 March 2011).

4. Harmelen, Frank van. European Research Consortium for Informatics and Mathematics. January 2008. http://ercim-news.ercim.eu/which-future-web (accessed on 5 April 2011).

5. Marshall, C.C. & Shipman, F.M. Spatial hypertext: Designing for change. *Communications ACM*, 1995, **38**(8), 88-97.

6. Greenberg, Jane. Knitting the semantic web. *Catalog. Classifi. Quart.,* 2007, **43**(3-4).

7. MusicBrainz. http://musicbrainz.org/ (accessed on 5 April 2011).

8. McGuinness, D.L. Ontologies come of age. *In* Spinning the semantic web: Bringing the world wide web to its full potential, edited by F. Dieter; J. Hendler; H. Lieberman & W. Wahlster. MIT Press, Cambridge, 2003. pp. 171-94.

9. RUSA guidelines for information services. http://www.ala.org/ala/mgrps/divs/rusa/resources/guidelines/guidelinesinformation.cfm (accessed on 17 April 2011).

10. Dhiman, Anil Kumar & Sharma, Hemant. Marketing mix in library and information centres. *In* ICAL 2009–Advocacy and Marketing. pp. 456-60.

11. World Wide Web. wikipedia.org/wiki/World_Wide_Web (accessed on 5 April 2011).

12. Building heterogeneous networks of digital libraries on the semantic web. http://www.slideshare.net/skruk/building-heterogeneous-networks-of-digital-libraries-on-the-semantic-web (accessed on 14 March 2011).